The MassGIS Protected and Recreational OpenSpace GeoDatabase Data Model

INTRODUCTION

MassGIS OpenSpace started around 1988 with the digitization of USGS Topo sheets. Many of these original features are still in OpenSpace (arc_date = 1901).

Coverage topology behavior can be modeled in a geodatabase by a combination of rules.

MassGIS uses ESRI's ArcSDE, with a 3rd party RDBMS (Oracle 9) for all production-level datasets, as this environment provides the efficient storage and high query speed necessary in high volume environments (e.g. multi-user environments or web services). SDE, with Oracle, uses the GeoDatabase data structure, implemented with ArcGIS 8.0. The GDB is the "workhorse" data structure for most ESRI products and was designed to have no practical limits to data size for storage (when implemented with ArcSDE), display (only fetches data relevant to the current view), or analysis (handles processing of large datasets).

The MassGIS Protected and Recreation Open Space Datalayer (hereafter referred to as OpenSpace) takes full advantage of this new data model by having custom GeoDatabase features, a fully relational set of tables to better model OpenSpace, custom tools for editing and analysis in ArcMap, better modeling of complex ownership/interest issues concerning real property, and feature topology. The "rules-based" topology of geodatabase is very different from the "forced" topology of ArcInfo coverages, the former data structure for OpenSpace. This provides an opportunity for dramatic improvement to our data structure and should serve OpenSpace in Massachusetts well for years to come.

Due to the fact that the transfer of data from a coverage to a geodatabase was such a large change in the structure of OpenSpace, it was implemented it in two phases. Phase I consisted of moving the "flat-file" model to ArcSDE. The first step involved a major cleaning of the dataset (fixing typos, etc.). OpenSpace was then be split into 3 feature classes; OPENSPACE_ARC, OPENSPACE_POLY, and CHAPTER61_POLY (the last since deleted) that all participate in a single topology in a OpenSpace feature dataset, mimicking the spatial integrity of a single coverage. Several new fields were added and defunct fields were dropped. Phase I is now complete.

Phase II was intended to take the last step of breaking the "flat" model into the fully relational model. It has since been determined that the query and mapping tools available in ArcGIS 9 do not support a full relational database structure for feature attributes. Therefore, there will be no Phase II as initially envisioned. However, these has been one additional major change to OpenSpace since PhaseI: the Chapter61 feature class has been removed entirely from MassGIS datasets. Chapter61 classification is in real life very dynamic and the MassGIS data layer had not been updated since the initial OpenSpace data collection effort the mid-1990's, which collected only incomplete Chapter61 data. The data have been removed to avoid its misused. Chapter61 data is now available through other datasets, including assessors' parcels.

HISTORY – MOVING FROM COVERAGE TO GEODATABASE

The first step was to examine the existing coverage data model and determine what should be kept, what should be dropped and what should be added. Looking at the feature classes involved in the coverage, and knowing that the labeling engine in ArcGIS is relatively robust, it was clear that the annotation layers could be dropped along with the regions (on site name) that existed only for the purpose of clearly labeling sites comprised of multiple parcels. The coverage annotation has become somewhat corrupt and does not appear worth the effort to migrate. In ArcGIS 9 we will be able to implement feature-linked annotation that will become the standard for OpenSpace in the future. This left us with labels, arcs and polygons. Without hard-wired coverage topology, the label attributes can be incorporated into the new polygon feature class. The arc feature class contains the very useful but oft ignored "code" field that has not been maintained very well over the years. Although this information could be dropped as the corresponding line segments of the polygons get snapped, it was determined that there was value to be had in retaining the arc feature class as an element to combine with the polygon feature class using a shared topology in a feature dataset. The arc layer is modified to now retain a history of source data and edits for each line segment (node to node). This will allow better estimation of the spatial quality of any given polygon in the new model.

The use of field domains will simplify editing and data maintenance.

An important feature of GeoDatabase technology is the ability to easily make attribute domains for simplified editing and constraining entered attributes for some fields. This will be apparent to the editor as drop down lists of valid attributes and validation of attributes. The domains provide a database enforced constraint upon what values may be entered into associated fields. ArcGIS will not allow the editor to enter data that falls out of the pre-defined domain bounds.

OSNAMES was the INFO table used in the coverage model to record all abbreviations used as OpenSpace field values. It was neglected for several years and subsequently many abbreviations were erroneously introduced into OpenSpace.

The biggest change though, is the breaking up of the coverage model into two related feature classes. Along with dropping some no longer needed fields, many new fields were added to better model land ownership and separated rights in land in Massachusetts. Arcs are now modeled as a separate feature class to allow tracking of individual arc edits/sources. Arc and polygon feature classes are tied together by a set of topology rules describing their defined spatial correlation. The OSNAMES table has also been explicitly incorporated into SDE to allow for better attribute coding via the custom OpenSpace edit form. The full description of the new attributes and tables follows.

THE NEW GEODATABASE DATAMODEL

Taking the former attributes tables for polygons and arcs, the overall usefulness of the existing model was evaluated and altered on paper through a series of meetings and discussions with OpenSpace editors. A few steps were completed before altering the data structure.

Data Preparation:

The first step was getting the 351 LIBRARIAN coverage tiles into a single statewide coverage. This step proved to be difficult as the resulting coverage could not be built due to intersection errors at town boundaries. Years of clipping coverage tiles to the town bounds had presumably resulted in a bit of rounding error (the data was single precision) and thus the tiles didn't fit anymore. To fix all these errors (many thousand) would take a long time of protracted editing. It was then decided to move the errors along into SDE and take advantage of versioned editing and rules-based topology in an SDE GeoDatabase. First, we could construct polygons from the old OpenSpace arcs and label points in SDE regardless of the poor topology. Since all OpenSpace polygons were complete and closed, the only errors are from overlaps between tiles thus ensuring that every polygon gets moved into the GeoDatabase. Arc8 topology was then generated on top of the data to find these intersections and overlaps. These will be fixed over time by the DBA whilst editors can continue to add new data in the versioned environment of SDE — again, a nice feature of a versioned GeoDatabase.

The next step was the most arduous — cleaning the attribute data. This made use of the frequency command (available for a GeoDatabase at Arc9) to find all instances within a field value. The results were then sorted and cleaned up in the statewide dataset. Not an easy task, but unlike the topology, this could be put on the versioned back-burner — we need to use the data itself in the new editing form (self-validation!) so it must be clean for everything to work properly.

The INFO table, OSNAMES has been cleaned up and expanded for use in the new edit form. This will be used as a self-referential lookup table for the various interest fields.

Commonwealth/EOEA interests have been cleaned up for use with the new edit tools and for proper symbolization on the FEESYM & INTSYM fields.

At present, no aliases have been established for OpenSpace analogous to the Alternate Name option for coverages.

Non-Profit interests were broken into three categories: Non-Profits (Type='N'), Land Trusts (Type='L') and Conservation Non-Profits (Type='G'). The land trust values were taken from the Massachusetts Association of Land Trusts listing available at http://www.massland.org/pages/neartrust/mainlist.html (Actually, there are a few more land trusts not affiliated with MassLand included in OpenSpace that are also coded 'L').

Actual holdings that are covered with water were reassessed. Previously, this was coded as SFO = 'W' and only used for MDC land. It has become apparent that this is a more

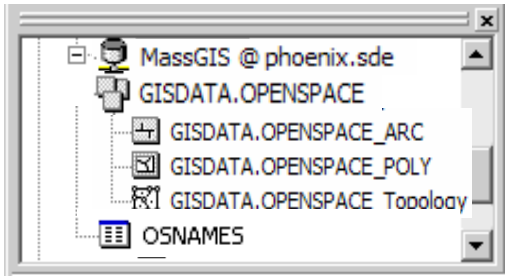
complex issue, and not only for DCRW. Therefore, these sites have been recoded to the owner's true type and have PRIMARY_PURPOSE set to 'U' for Underwater.

Municipal interests were reconciled with OSNAMES and the town itself, although this work continues (What are the departments really called vs. what did the volunteer tell us?).

Additional fixes are being done in the background by the DBA during active editing by agency OpenSpace Editors.

THE NEW GEODATABASE DATAMODEL SCHEMA:

OpenSpace Feature Dataset



OpenSpace Feature Dataset

In the root level of the MassGIS SDE GeoDatabase lies the new GISDATA.OPENSPACE Feature Dataset. Inside this dataset are the 2 OpenSpace Feature Classes and the Topology Feature (which is treated as separate feature class by ArcGIS). Outside of the Feature Dataset lies the OSNAMES table required for using the custom edit form (tables cannot be included in a feature dataset).

GISDATA.OPENSPACE_POLY

This is the primary feature class for Open Space. There are several new fields as noted below and some fields that have been dropped. The table to the right is a brief description of the feature class attribute fields.

The SCORP_ID field has been deleted as it is about 15 years out-of-date. A table remains in SDE that links OS_ID to SCORP_ID if there is ever a need for that record in the future. The POLY_DATE field has been dropped – the date of last spatial alteration for a parcel now exists in the ARC feature class. This will yield finer grained spatial edit attribution. The COUNTY_CODE field has also been dropped as counties are no longer legal entities. Relic coverage fields have also been dropped (including many redefined fields).

Many of the new fields were added to resolve existing or foreseen limitations in the existing data model. In the following section is a description of each field and its intended use.

TOWN_ID	A 3 digit integer uniquely identifying every municipality in Massachusetts. The values in this field range from 1 (Abington) to 351 (Yarmouth). **[Domained]**
POLY_ID	A 5 digit integer unique identifying every feature in the given municipality. This field needs to remain unique relative to the parcel of land rather than to ArcGIS. For this reason we do not use the ArcGIS generated OBJECTID.
SITE_NAME	The name associated with the parcel, if any (e.g. Jones Park).
FEE_OWNER	The name of the holder of the deed to the land or the grantor of the land represented in the polygon.
ABRV_FEE_OWNER	The link to the OSNAMES table. A simple and unique abbreviation for some of the lengthier names in the field.
STATUS_FEE_OWNER	Category for the fee owner's status. **[Domained]**
MANAGER	The name of the entity that maintains the property if different from the fee owner.
ABRV_MANAGER	The link to the OSNAMES table.
STATUS_MANAGER	Category for the manager's status. **[Domained]**

OPENSPACE POLY schema

Simple feature class
OPENSPACE_POLY

Geometry *Polygon*
Contains M values *No*
Contains Z values *No*

Field name	Data type	Domain	Comments
OBJECTID	Object ID		internal field
Shape	Geometry		internal field
TOWN_ID	Short integer	Town_ID1	Range: 1 - 351
POLY_ID	Long integer		Limited to 99999
SITE_NAME	Text - 120		
FEE_OWNER	Text - 100		
OWNER_ABRV	Text - 20		Coded from OS_NAMES
OWNER_TYPE	Text - 1	OS_Type	Coded
MANAGER	Text - 100		
MANAGER_ABRV	Text - 20		Coded from OS_NAMES
MANAGER_TYPE	Tex - 1	OS_Type	Coded
PRIMARY_PURP	Text - 1	OS_Primary_Purpose	Coded
PUB_ACCESS	Text - 1	OS_Public_Access	Coded
LEV_PROT	Text - 1	OS_Level_Protection	Coded
OLI_1_ORG	Text - 100		
OLI_1_ABRV	Text - 20		Coded from OS_NAMES table
OLI_1_TYPE	Text - 1	OS_Type	Coded
OLI_1_INT	Text - 20	OS_Interests	Coded
OLI_2_ORG	Text - 100		
OLI_2_ABRV	Text - 20		Coded from OS_NAMES table
OLI_2_TYPE	Text - 1	OS_Type	Coded
OLI_2_INT	Text - 20	OS_Interests	Coded
OLI_3_ORG	Text - 100		
OLI_3_ABRV	Text - 20		Coded from OS_NAMES table
OLI_3_TYPE	Text - 1	OS_Type	Coded
OLI_3_INT	Text - 20	OS_Interests	Coded
GRANTPROG1	Text - 20	OS_Projects	Coded
GRANTTYPE1	Text - 1	OS_Type	Coded
GRANTPROG2	Text - 20	OS_Projects	Coded
GRANTTYPE2	Text - 1	OS_Type	Coded
PROJ_ID1	Text - 20		State use only
PROJ_ID2	Text - 20		State use only
PROJ_ID3	Text - 20		State use only
EOEAINVOLV	Short integer	OS_EOEA_Inv	Coded
ARTICLE97	Short integer	OS_Art97	Coded yes/no
FY_FUNDING	Short integer	FiscalYear	Range: 1800 - 2010
BOND_ACCT	Text - 20		State use only
GIS_ACRES	Double		Calculated field
CAL_DATE_REC	Text - 10		Use Jan. 1 if year is unknown
DEED_ACRES	Double		
OS_DEED_BOOK	Long integer		
OS_DEED_PAGE	Long integer		
ASSESS_ACRES	Double		
ASSESS_MAP	Text - 10		
ASSESS_BLOCK	Text - 10		
ASSESS_LOT	Text - 10		
ASSESS_SUBLOT	Text - 10		
ALT_SITE_NAME	Text - 120		
ATT_DATE	Date		Calculated field
BASE_MAP	Text - 10		
SOURCE_MAP	Text - 50		Derelict field
SOURCE_TYPE	Text - 10	OS_Source	Coded
COMMENTS	Text - 255		
LOC_ID	Text - 15		
DCAM_ID	Long integer		
FEESYM	Text - 20		Calculated field
INTSYM	Text - 20		Calculated field
OS_ID	Text - 8	Unique ID !	Calculated field
Shape_Length	Double		internal field
Shape_Area	Double		internal field

PRIMARY_PURP — A single character text code indicating the *initial* reason the land was acquired as open space. In most cases, this is also the current use of the land, but there are some exceptions. The most frequent exception is municipal land acquired many years ago for water supply that has then been discontinued as a public water supply and converted into recreation/conservation land. **[Domained]**

PUB_ACCESS — A single character text code indicating the *legal* level of public access (not to be confused with physical access such as street frontage). For most parcels, public access is either open (public is welcome on the parcel) or closed (no public allowed). **[Domained]**.

LEV_PROT — A single character text code indicating the relative impediment to the parcel being developed. As no parcel of land can ever be "permanently" protected, we consider many different types of land interest to impart protection in perpetuity. These include Article 97 lands (e.g. EOEA agency land), non-term Conservation Restrictions, land held by land trusts and environmental non-profits, etc. All parcels are sorted into level categories on a parcel-by-parcel basis. **[Domained]**

OLI_1_ORG — The name of the holder of the associated interest.

OLI_!_ABRV — The link to the OSNAMES table. A simple and unique abbreviation for some of the lengthier names in the field.

OLI_!_TYPE — Category for the interest holder's status. **[Domained]**.

OLI_1_INT — The type of interest held by OLI_1_ORG. **[Domained]**. <<2nd and 3rd interests>> *Ibid.*

GRANTPROG1 — Grant program associated with the parcel. **[Domained]**

GRANTSTAT1 — Category for the grant status. **[Domained]**. <<2nd grant>> *Ibid.*

PROJ_ID1 — Project identifier for the parcel; may include many parcels under a single project (EOEA use only). <<2nd and 3rd projects>> *Ibid.*

EOEAINVOLV — Indicates the category of funding the parcel received from EOEA sources. **[Domained]**. (EOEA use only).

ARTICLE_97 — Yes/No. Is the parcel protected under article 97 of the Massachusetts Constitution? **[Domained]**. (EOEA use only).

FY_FUNDING — Fiscal year project was completed (EOEA use only). **[Domained]**.

BOND_ACCT — Funding of parcel acquisition/protection (EOEA use only).

GIS_ACRES — GIS Calculated acreage of parcel.

CALDATEREC — Calendar date deed was recorded. If only year is known, it is set to January 1st of that year.

DEED_ACRES — Acreage according to the recorded deed.

OSDEEDBOOK — The number of the book the deed for this parcel and/or interest was recorded in at the local registry of deeds. **[Domained]**.

OSDEEDPAGE — The starting page of the above recorded deed. **[Domained]**

ASSESS_ACRES — Acreage according to the local assessor maps or database.

ASSESS_MAP — Tax map identifier text as determined by the Assessor

ASSESS_BLOCK — Tax map block identifier as determined by the Assessor

ASSESS_LOT	Tax map lot identifier as determined by the Assessor
ASSESS_SUBLOT	Tax map sublot identifier as determined by the Assessor
ALT_SITE_NAME	A second name associated with the parcel, if any (e.g. No Town State Forest).
BASE_MAP	Number of the MassGIS basemap as indicated in the upper right corner of the map. This is the map data was recompiled upon. Only used for volunteer non-digital updating. (EOEA use only).
SOURCE_MAP	Code linking to the Source Map Worksheet delineating the specifications of the map that the polygon information was taken from.
SOURCE_TYPE	Text code indicating what the source data was to give a better estimation of the quality of the polygon attributes. This will be used for the entire polygon. For more specific source information on the arcs comprising the polygon, see the OPENSPACE_ARC feature class. **[Domained]**.
COMMENTS	No comment.
ATT_DATE	Date of last attribute edit.
LOC_ID	12 digit number identifying the coordinates of the centroid of the parcel in MassGIS Standard Parcels.
DCAM_ID	7 digit integer for linking to the DCAM data tables via the ARC_ID field.
FEESYM	Field used for symbolization of ownership (EOEA edits only).
INTSYM	Field used for symbolization of separated rights to OpenSpace land (EOEA edits only).
OS_ID	This is the unique statewide identifier taken from the coverage model. It has been changed from the old 7 digit integer where the first 3 digits (including leading zeroes) are the TOWN_ID number for the town the polygon exists in and the last 4 digits are the unique identifier for that town (formerly POLY_ID). As editing over the years has produced many various POLY_ID values, it was necessary to bump POLY_ID to a 5 digit integer. This has forced OS_ID to be altered to a 9 character string field. The first 3 characters are the TOWN_ID as before immediately followed by a dash. The last 5 characters are the POLY_ID. This archaic structure is a relic from the early coverage days of OpenSpace, but is retained as it links tens of thousands of pages of source documentation to the database. This field populated by the DBA.

GISDATA.OPENSPACE_ARC

OPENSPACE ARC schema

Every polygon in the OPENSPACE_POLY Feature Class will have a boundary covered by OPENSPACE_ARC. The arcs will contain more information that they have previously. The old POLY_DATE field is now stored in the arcs due to the piece-meal editing/ updating of many parcels. We need to retain what portion of a parcel was edited when and with what source data.

Simple feature class
OS_ARC

Geometry Polyline
Contains M values No
Contains Z values No

Field name	Data type	Domain	Precision	Scale	Length
OBJECTID	Object ID				
Shape	Geometry				
CODE	Short integer	OS_Arc_Code	2		
SOURCE_TYPE	Text	OS_Source			10
ARC_DATE	Date		0	0	8
COMMENTS	Text				100
DROPLINE	Text				10
ARCID	Long Integer		7		
Shape_Length	Double		0	0	

Akin to the ArcEdit coverage edit model, we can take points and arcs and create polygons (explicitly avoiding the term 'Build'). Also, for ArcView 3x editors, we can also create arcs from polygons using Map Topology (different from coverage topology and rules-based topology!). With these caveats and a detailed custom edit form, we can enforce a topological rule between OPENSPACE_ARC and the two polygon feature classes that all polys must be bounded by arcs. More detail on this process is in the OpenSpace SDE Editing Manual.

The Arc feature class has fields as follows:

CODE	This is the old code field from LIBRARIAN that links OS features to physical features
SOURCE_TYPE_PRIMARY	Text code indicating what primary source data was to give a better estimation of the quality of an individual arc.

For use when a polygon is comprised of arcs from sundry sources. **[Domained]**.

SOURCE_TYPE_SECONDARY Text code indicating what secondary source data was to give a better estimation of the quality of an individual arc. For use when a polygon is comprised of arcs from sundry sources. **[Domained]**.

ARC_DATE Date of last spatial revision to arc in question. Assumes role of POLY_DATE.

COMMENTS No comment.

DROP_LINE Code to enable site mapping by not drawing the internal arcs.

GISDATA.OPENSPACE_TOPOLOGY

The rules-based GeoDatabase topology appears as a feature class in the Open Space Feature Dataset. Unlike coverage topology, the GeoDatabase topology allows for errors to persist without affecting the rest of the database. The rules used are defined as follows:

I.	OPENSPACE_ARC	Must not Overlap	
II.	OPENSPACE_ARC	Must not Self-Overlap	
III.	OPENSPACE_ARC	Must not Have Dangles	
IV.	OPENSPACE_ARC	Must be Single Part	
V	OPENSPACE_ARC	Must not Self-Intersect	
VI.	OPENSPACE_ARC	Must Cover Boundary of	OPENSPACE_POLY
VII.	OPENSPACE_POLY	Must not Overlap	
VIII.	OPENSPACE_POLY	Boundary Must be Covered by	OPENSPACE_ARC

This set of rules approximates coverage topology. The editing tools of ArcGIS allow for the editor to create polygons from arcs or create arcs from polygons – this allows the editor to use whichever method they prefer for creating new features. They then use the automated tools to create the associated features required by the topology.

GISDATA.OSNAMES

OSNAMES is a table outside of the OpenSpace feature dataset that lists all abbreviations used in OpenSpace. It is used for populating the OpenSpace attribute editing form.

OSNAMES schema

Table
OSNAMES

Field name	Data type	Domain	Comments
OBJECTID	Object ID		internal field
AGENCY	Text - 100		
ABBREV	Text - 20		
DEPT	Text -8		for data mangement
TYPE	Text -1	OS_Type	coded

AGENCY Full text of entity's name

ABBREV Abbreviation of entity's name

DEPT Internal table reference between the FEE_OWNER and MANAGER fields in the polygon attribute table.

TYPE Code specifying the classification of entity..

Epilogue

That's the general idea behind the new OpenSpace data model. Please contact Ben Smith, OpenSpace Database Administrator; with any questions or concerns regarding OpenSpace (benjamin.smith@state.ma.us, 617.626.1076). Updates to lands owned and managed by municipal and non-profit organizations are very welcome, particularly in digital form (shapefiles are fine); please be aware these updates may not appear in OpenSpace immediately. Be sure to check out the OpenSpace Website http://www.mass.gov/mgis/osp.htm .

Scott Castello
former OpenSpace DBA
MassGIS / EOEA